

Identification and characterization of new human medium reiteration frequency repeats

Jerzy Jurka, David J.Kaplan¹, Craig H.Duncan¹, Jolanta Walichiewicz, Aleksandar Milosavljevic, Gayathri Murali¹ and Joseph F.Solus¹

Linus Pauling Institute of Science and Medicine, 440 Page Mill Road, Palo Alto, CA 94306 and

¹Center for Molecular Biology, Wayne State University, Detroit, MI 48201, USA

Received October 14, 1992; Revised and Accepted January 29, 1993

ABSTRACT

We report nine new families of human medium reiteration frequency interspersed repetitive elements (MER elements). They were identified by computer-assisted analyses. Six of them were independently confirmed as repetitive families by DNA–DNA hybridization, and the number of elements for each of these families was estimated by plaque hybridization assay. The involvement of some of the reported MER elements in genetic rearrangements is demonstrated.

INTRODUCTION

The current number of known human repetitive families exceeds fifty (1). Most of these families represent medium reiteration frequency (MER) repeats of unknown origin with a copy number per genome ranging from hundreds to thousands. These repeats represent a fossil record of evolutionary processes shaping the human genome and affecting chromosomal stability of the contemporary human individuals (1). Therefore, studies of repetitive DNA are of both biological and medical importance. MER repeats are also of practical significance for physical mapping of the human genome (2).

Nine new MER families are identified by computer-assisted analyses of the GenBank DNA database, and six of them are characterized experimentally in this paper. Together with previous work (3,4), this paper further illustrates effectiveness of the combination of computer-assisted studies and experimental approaches.

MATERIALS AND METHODS

Known repetitive sequences have been eliminated from GenBank DNA (Release 69.0). The processed DNA sequences were systematically compared and aligned with each other as previously described (3). The alignments between seemingly unrelated sequences were studied in the context of the original literature, and prospective repetitive elements were selected for hybridization studies. The list of GenBank loci names and definitions of sequences containing newly identified MER repeats are given in Table I.

The MER sequences used in hybridizations are shaded in Table III. MER sequence probes were provided by Lagan Inc. (Detroit

MI) as purified double stranded DNA. Double stranded DNA was labeled by the random primer method (5). Hybridizations to filter-bound DNA(6,7) were performed at 60°C in a commercial hybridization mix (Amersham Corp.) for 16 hrs. Washing was done twice for 5 minutes in 2×SSC, 0.1% SDS at 30°C, followed by two washes for 15 minutes in 2×SSC, 0.1% SDS at 50°C. Autoradiography was performed in the presence of an intensifying screen at –70°C. The repetition frequencies for individual MER families were determined by the plaque hybridization assay (4).

RESULTS

Characterization of MER families

MER25. This repeat was first characterized by Rogan et al. (8) as a potentially distinct repetitive family. It was confirmed by the additional two sequence examples of MER25 which we identified in GenBank DNA, and by plaque hybridization assay (4). Based on the assay the number of MER25 repeats in the human genome is estimated to be around 2500. All estimates for individual families given below were done using the same method.

The MER25 sequences listed in Table III are 76.9% similar, and the longest one (HUMHBB) is 453 bp long. One of the MER25 repeats represents a polymorphic marker closely linked to the cystic fibrosis locus (13). Hybridization of MER25 to mammalian genomic blots (Figure 1) demonstrates the presence of the sequence on a disperse set of DNA fragments in both human and monkey DNA. As is the case for many of the previously described MER sequences (4), there is a pattern of discrete bands imposed upon the polydisperse background.

MER26. Six examples of this repeat have been identified in GenBank, and the longest sequence (HUMIGLAMB) is 138 bp long. Due to the short length and low average similarity (66.1%) the DNA–DNA hybridization signals are relatively weak, and we were unable to obtain reliable overall frequency of these repeats.

MER27. Four examples of this repeat have been identified. Their average similarity is 69.2% and the longest (HUMSIGMG3) is 170 bp long. This sequence is followed by MER11 reported in

Table I. GenBank loci names and definitions of sequences containing new MER repeats

LOCUS NAME	MER	SOURCE SEQUENCE	REF.
HUMABL9I	30	c-abl oncogene intron.	34
HUMABLI2	30	ABL region of intron 1a/a2 involved in Ph translocation.	35
HUMAFP	32	alpha-fetoprotein gene, complete cds (homologous to gorilla: GORAFPA ref.41).	42
HUMAMINONA	26	aminopeptidase N gene, exon 1.	21
HUMAPB03	27	apolipoprotein B-100 (apoB) gene, exon 4.	26
HUMAPOAI	26	apolipoprotein A-I (ApoA-I) gene, exon 1.	22
HUMBFXIII	25, 26	factor XIII b subunit gene, complete cds.	20
HUMCDIR2	32	CD1 R2 gene for MHC-related antigen.	43
HUMCGPRA	30	cGMP phosphodiesterase alpha subunit (CGPR-A) mRNA, complete cds.	36
HUMCRYGBC	29	gamma-B-crystallin (gamma 1-2) and gamma-C-crystallin (gamma 2-1) genes, complete cds.	31
HUMDAFC5	32	decay-accelerating factor (DAF) gene, intron 2, partial.	44
HUMDMDH7	31	Duchenne muscular dystrophy DNA fragment (overlaps 1.2 and 3.8 kb Hind III fragments) with exon Y.	17
HUMFVIII12	33	clotting factor VIII deletion junction.	16
HUMGP91	29	GP91-PHOX gene promoter region.	12
HUMHBB	25, 32	beta globin region on chromosome 11.	19, 45
HUMHBEG	25, 27	L1Heg repetitive element from the intergenic region of the epsilon and G-gamma globin genes.	8
HUMHPRTB	30	hypoxanthine phosphoribosyltransferase (HPRT) gene, complete cds.	37
HUMHS5R	28	specific HS5 DNA.	14
HUMIGLAMB	26	lambda-immunoglobulin constant region complex (germline).	24
HUMIL2RBA	27	interleukin-2 receptor beta chain gene.	9
HUMINTB1A	30	integrin beta-1 subunit mRNA, 3' end (cytoplasmic domain).	38
HUMKM19	25	KM19 gene.	13
HUMMLC1G5	28	MLC1emb gene for embryonic myosin alkaline light chain, 3' UTR.	15
HUMPGEF102	29	snRNP E protein pseudogene 110.	32
HUMPHLAM	28	phospholamban mRNA, complete cds.	29
HUMPLG2BA	32	platelet glycoprotein IIb mRNA, 3' end.	46
HUMRSKP2	29	KpnI repeat 1.8 kb family member DNA flanking region interrupting a nuclear sequence homologous to mtDNA.	33
HUMSIGMG3	27	sigma(gamma)3 DNA upstream of Ig C(gamma)3 gene (sigma G3).	28
HUMSTXG2	31	chromosome X steroid sulfatase (STS-X) gene, exon 10 (identical with HUMSTS).	18
HUMTATG5	31	gene for tyrosine aminotransferase (TAT) (EC 2.6.1.5) 5'- flanking region.	40
HUMTFPB	28, 33	tissue factor gene, complete cds.	30
HUMTPO04	32	thyroid peroxidase (TPO) gene, exon 4.	23
HUMUKI3	32	urokinase inhibitor (PAI-2) gene, exons 3-5.	27
HUMXT00121	26	expressed sequence tag (EST00121).	25
HUMXT00273	26	expressed sequence tag (EST00273 similar to Human ApoAI 5' untranslated).	25

Table II. Summary characteristics of MER repeats

MER	No. of copies per GenBank	Similarity [%]	Repetition Frequency		Max. obs. length [bp]
			observed	estimated	
25	3	76.9	2500	1000	453
26	6	66.1	-	2000	138
27	4	69.2	-	1333	170
28	4	73.2	2500	1333	403
29	4	72.6	4000	1333	498
30	4	81.6	-	1333	210
31	3	71.6	500	1000	163
32	7	74.0	2000	2333	154
33	2	88.9	-	667	81

From left to right: (1) names of MER sequences; (2) numbers of MER repeats identified in GenBank release 69.0; (3) average pairwise similarity; (4) observed and estimated repetition frequencies, and (5) maximal observed lengths.

the previous paper (4). It cannot be excluded that MER27 and MER11 represent different portions of the same repeat. Similarities between HUMIL2RBA and other loci have been identified before (9), but no biological explanation was given. This family was not studied experimentally.

MER28. Four elements from this family are listed in Table III, and the most complete sequence (HUMMLC1G5) is 403 bp long. Overall sequence similarity of this sample is 73.2%. This family

contains around 2500 sequences in the human haploid genome. These sequences are present on a disperse set of restriction fragments in digests of human and monkey DNA (Figure 1).

Fragments HUMTFB;1 and HUMTFB;2 most likely originated from a single MER28 repeat, which did undergo some internal duplication followed by an insertion of Alu-Sx (10). Alu-Sx is among the oldest Alu sequences. This information combined with the 73.2 sequence similarity in MER28 indicates that this is an old family, probably older than the Alu-Sx subfamily (10). The

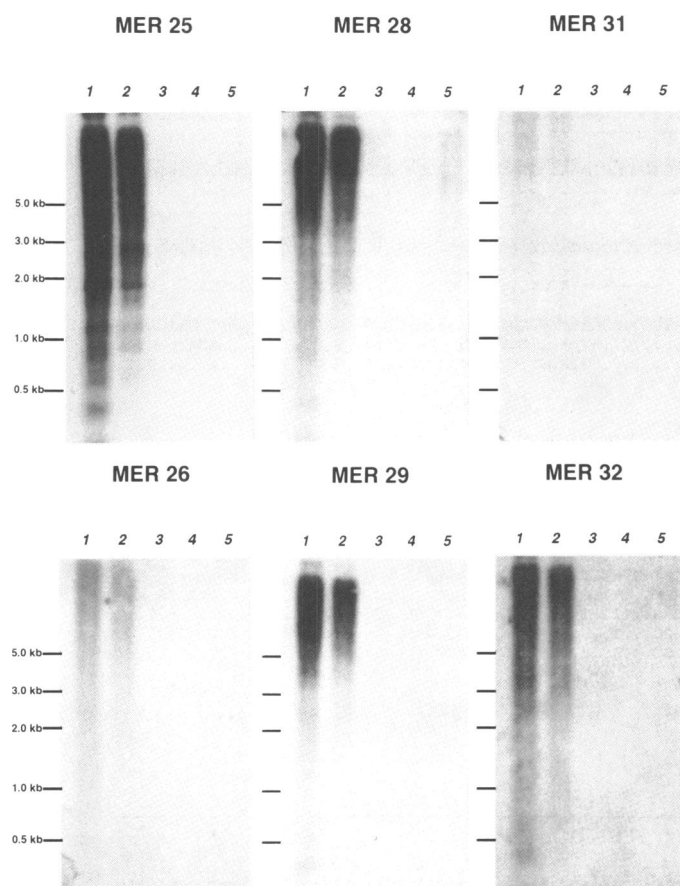


Figure 1. Hybridization of mammalian chromosomal DNAs with MER DNA probes. Five μ g samples of EcoRI restriction endonuclease digested chromosomal DNAs were subjected to electrophoresis through a 0.7% agarose gel and transferred to nylon membranes. The indicated lanes contain the following mammalian DNAs: (1) human, (2) Rhesus monkey, (3) mouse, (4) Chinese hamster, (5) cow. Autoradiographs of the membranes hybridized with radiolabeled MER sequences are shown. Positions of molecular weight markers are indicated by dashes at the right of each autoradiograph.

age of Alu-Sx is comparable with the age of Alu-J estimated to be 55 million years (11). Two MER28 sequences (HUMMLC1G5 and HUMTTFPB;1) are followed by fragments of MER8 repeat (ref. 4, marked by asterisks in Fig. 1). It remains to be determined whether this association is accidental or MER28 and MER8 are different parts of the same repetitive element.

One of the MER28 sequences (HUMHSSR) is within a human-specific sequence not present in chimpanzees (14). As MER28 appears to be an ancient family of repeats, we propose that it was inserted prior to the evolutionary separation of humans and chimpanzees and deleted after separation of these two species. Another MER28 (HUMMLC1G5) introduced alternative splicing site to the human embryonic myosin alkali light chain gene (15). As a result, an internal portion of this repeat is spliced out as a non-translated exon 8 (marked by lowercase letters in Table III).

MER29. Average similarity among the four identified sequences from Table III is 72.6%. The two longest repeats (HUMPGEP102 and HUMGP91) are 498 and 486 bp long, respectively. HUMPGEP102 is split approximately in the middle by an Alu-Sx insertion. As in the case of MER28 (see above)

this can be taken as an indicator of its age. Similarity between HUMCRYGBC and HUMGP91 was first detected by Skalniak et al. (12), but the authors did not provide any biological interpretation. This family is relatively large as it contains 4000 elements. These sequences are present in a disperse set of restriction fragments in digests from human and monkey DNA (Figure 1).

MER30. The four sequence examples presented in Table III are on average 81.6% similar which makes them the second most similar among sequences presented in this paper (after MER33). Their upper length is 210 bp (HUMINTB1A). The repeats are most variable in the central AT-rich region (see Table III). One of them (HUMABLIA2) contains an Alu-Sb insert. Overall, Alu-Sb represents the youngest known large Alu subfamily (10), although at least one branch within this subfamily called HS or PV (47, 48) seems to be even younger. The Alu-Sb insert does not belong to HS/PV branch, and it is not expected to be a very recent one. Based on the sample similarity and on the presence of the relatively young Alu insert, we tentatively conclude that MER30 may be the youngest among families presented in this paper. This family was not studied experimentally.

MER31. This family is represented by three sequences 71.6% similar and with maximum length 163 bp (HUMTATG5). MER31(HUMDMDHI7) is flanking a deletion-prone exon from Duchenne muscular dystrophy gene (17). Interestingly, one of the designed PCR primers is placed within this repetitive element (probe 8 from ref. 17). In addition, this repeat is a portion of a larger inverted repeat flanking both ends of the human X-linked steroid sulphatase gene (18). This family is relatively small as it contains ~500 elements and, predictably, the hybridization signal is rather weak (Figure 1).

MER32. This family is represented by 7 human sequences 74% similar; the longest sequence is 154 bp long. The overall number of MER32 elements in the human genome is around 2000.

MER33. The sequence immediately adjacent to the 3'-end of MER28+MER8 repeat complex from HUMTFFPB (Table III) is homologous at positions 8990–9070 to one of the deletion junctions in the human factor VIII gene (HUMFVIII12, position 337–477 ref. 16) obtained from patients with hemophilia A. This sequence may represent still another medium frequency interspersed repetitive family, which we tentatively call MER33. The breakpoint is within the putative MER33 repeat (position 470). This sequence was not studied experimentally.

Genomic abundance and detectability of MER repeats

The total number of MER25-MER33 sequences present in GenBank (release 69.0) is 37. This release contains around 13.5 million bp of DNA sequences, which represent 0.45% of the genomic DNA. A lower figure (0.3%) is more realistic, as it accounts for identical clones and cDNA. The 37 MER elements in 0.3% of the genome represented by the GenBank correspond to 12333 elements in the total genome (around 1400 elements per family; individual estimates are listed in Table II). The studies of five MER families (MER25,28,29,31 and 32) by the plaque hybridization assay gave average number around 2300 elements per family.

Estimates based on the small numbers of individual repeats in GenBank (see Table II) are rather uncertain, but the overall

Table III. Alignment of MER25-33 and their locations at sequenced loci

[illegible]

MER26

```
HUMIGLAMB GAAACGAATCTCC-CCAGAGTCCCAGA-AGGAGC-CG--GCCTGCTGGGACCCTGATTITAG-CCC-CCAAGGACTCATTTTGACTTCTGACCT--CCAGA
HUMBFXIII T A C T C A G C A C C A C C A G A A G C C C G . . . . . G . . . . T . G
HUMAMINONA . . . AGG . . . - TTCC . C . T . A . . . . - T . TAT . . . . AACCC . T . C . . . . T . A - G . G . . . . C . . . . . TT . TA . A . A .
HUMXT00273 . . . AT . . . --GT . . . C . T . . . . A . A . A - CGT . . . ACAC . TT . T . . . . T . - AGTGAA . . . G . . . . A . . . . . - T . .
HUMXT00121 . . . AG . . . T . T . A . . TG . G . . CT . - . A . . AC . T . . . . KT . - . AG . GAA . . . . . T . . . . G . . - C . .
HUMAPOAI . . . - TGC . C . T . - . . . A . A - . T . . CC . A . . . . . - AGT . T . C . . . . . T . . . - . T .
```

HUMIGLAMB CCTG--GGAGAGAGTCAGTGTGTGTATGTTTTAAGGCACTAAATT> 32880
HUMBFXIII A T A A A C C C C < 11792 (GTATTTCATCAAGCAG)
HUMAMINONA T . . -TA . . T . . A . A . TA . . T T . < 130
HUMXT00273 A . . -TA . . T . . -A . TCC . T C . < 196
HUMXT00121 AAA -TA . G . G . A . - . . C T . T . < 286
HUMAPOAI A . . -TA . . T . A . A . A . T . . ATC > 41

```

MER27
HUMSIGMG3 TAAAAAATACAGTCAGGTAGGAGGAATAGGCCCGGTGTTCTCTAACACCATAGGGTGACTATACTAACACCAATTTATTGTATATTTTCATACGGCTAGAAG
HUMIL2RBA ..CG...AC.G.A.T..A.C.A.T....A.AT.TA..A.TGA..G..TA.C.....T..GTC....AT.....C....AA.A.TAA...A..
HUMHBEG                                     .....GT...TTA.T.....A.....CA.A.T..TC....A

HUMSIGMG3 AGCAAATTTTGATTGTTCCCGAGCACAAAGAAGTGATAAATGTTTGAGGTGATGGATATGCTAAAT< 401
HUMIL2RBA ..T.T.C..G.....---A.....AG.....AC.....< 236
HUMHBEG T.AGG....GA.A.T..T..AT..T..A.....C.CA.A.A.....C....A.> 1484
HUMAPB03 ..AA.....A.....A.....G.....> 2131

```

[illegible]

MER29

HUMGP102 AGATTATAATGAAGCTGAAAAGTTCTGTCTCCTG-GTG---GGTATGCTATAGCTGTTCAT---CATG-GCAT--TAACCCAAACACATTTACTCACGTGTT-----
HUMGP91 .T.A.....AT.TAAT.T.TA.A.T....G.-G..T....-C...-AAA TC AC CTAB A TG G AT T TGCG
HUMRSKP2G...C.....A.C..CA.TG...A...ACTT.TA.T....C...A...AGC...GAA..TG..G.A...G.G.GGG.GTG...A-----

HUMGP102 -----TAATGTAAACAAACCTACTCGGCTACCAGTTGTATAAAAACTAGCACATAGAATTATGAATAGTAGACATAAATACTTGATAATCGATAATAAATGACTAT
HUMGP91 CAATOC GO G A C T C G A A B
HUMRSKP2 ---AC.GG..G.....C.....T...G...CC...C...G.A...A...C...G...T.C.....G-A-A...B-----

HUMGP102 TATCATAATA-TAGTAGTTGTGACTAAATTTATGTACAATTCAAATATA-TGTTTATACGT-ATTTTAGAGTATACxTCCTTCTACTTA-----TT-AAAAAGCA
HUMGP91 G C G T C T G C G C A C A C A
HUMRSKP2 ----.T..C.G.....GT.....G.....T...-GA...-T...C...G...-T.....TA.
HUMCRYGBC C.C.T.T -A.C.C.C.T T TACTTTT A AA

HUMPGEPI02 AAAGTTAACTATAAAACAGCCCCAGGCAGGTCCATCAGGAGATATTTTCAG-AGGAAGGCATTGCTATCATTACAGATAGACGTCCATGTGTGTTTCTGCCCTGA
 HUMGP91G.....G.AT.T.....T.....A.G.....C.....A.T.....G.....GAT.C.....-T.....CA..AG.....T.....
 HUMRSKP2C.....< 573
 HUMCRYGBCG.....CAN..TT..-..AG..T.....G..C.A..A.....-..CATC.....GGT..C..-T.....T.....G.....

HUMPGEPI02 AGACCTTCCAGTGGGACAAGATGTGGAGATGCAAGACAGTGATATTAATGATCCTGACCTGGGTGGGCCTAGGATAATGTGTG----TGTGTTTTCAGCTTTTAA
 HUMGP91TG.....G.....A.....G.....G.....TT.T.....C.....ATGTT.....C.....
 HUMCRYGBCT.....T.....C.....AG..GT.....G.....C.T.A.....AAC.....TGTT.....C.....T.....C.....

HUMPGEPI02 CAAAAAAGT< 4485
 HUMGP91T..> 55
 HUMCRYGBC< 8288

MER30

HUMINTB1A GTATATAGCAGGGGTGCCAATCTTTTGGCTTCCCTGGGCCACACTGGCAAGAAGAAGAACTGCTTGGGCCACACATAAAAT-----ACACTAACGATAGCTG
 HUMHPRTBCC.....A.....A.....T.....G.....AC..A..-
 HUMABLIA2A.....-.....T.....A.T.....C.....C.....T.....
 HUMCGPRAAC..T.....T.....T.....T.....ACAGTA.....T.....

HUMINTB1A ATGAGCTTTAAAAAAGTTACCAAAAAAATAAT-AATAATAATTA---CAATTTCT-AAGAAAGTTTACAAATTTGTGTTGGGGCCACATTCAAAGCTATTCT
 HUMHPRTB-G.T..C.....-C.A..A..TCATG.TG..T.....T.....G..T.....C..C.C
 HUMABLIA2-.....C.....C.....x..A..ATC..-TA.TG..T.T.....G.....-A.....< 522
 HUMCGPRA T..A..< 2889

HUMINTB1A GGGTGCATGT> 524
 HUMHPRTBTC...G..< 45902

MER31

HUMTATG5 TCTTGGCATAAAATCCACA-CTTGCCA-T-CACCATATTTGGAGTTGAATCCAAT-T-TCTCT-TCC--CC-ACGGAA-AATCCATTCAGTGGGCGCTAG
 HUMSTXG2C..T.....-GGT.....CA...C...G...G..C.....C.....T...AC..GTCC.....T.....T...CCG..C
 HUMDMHI7T..CT.....C..GTG..TG.....A.....GC.....C.C.....A..AA..TC...T..T.C.....T..C.G.GC

HUMTATG5 CTATTGCAATGGTCTGAAATAAGTCTGCCTTACCATGCTTTAACAAGTGTATTGAATAATTTTTT< 382
 HUMSTXG2 ..G..CATG.....A.....-..TG.....T.G..AGC.....A.....< 6144
 HUMDMHI7TGG..TA.....T.....CA.....AA.....> 490

MER32

HUMPLG2BA AAAATTTATGTATTTCCAGCTGTATTGAGGTATAATTGACAAAT--AAATTGTATATATTCACAGTGTGCAATATGATGTTTGGTATATGTATACATTGTGAAA
 HUMDAFC5-..CC.G--..T.....C.....A.....-C.T-AG...A..C.....A.....GC.....
 HUMHBBA..T.....C..C.....AA.....T..AG...A..C.....AA.....C.....
 HUMCDIR2CCTCCC.G.....T..C.....G.....AA.....C..G...AGA..A..G..G...AC..C..CC.....
 HUMUKI3ACC-T..T.....T.....C..G.....AC.....A.....G..T.GG..AA..-G.....CAA.....T.....
 HUMAFPACC-T..T.....T.....C..G.....AC.....A.....G..T.GG..AA..-G.....CAA.....T.....
 GORAFPAACC-T..T.....T.....C..G.....AC.....A.....G..T.GG..AA..-G.....CAA.....T.....
 HUMTP004-..GA.....A.....A.....CAA.....AA.....

HUMPLG2BA GGATTACCACAATCAAGCTAA-TTAGCGTATGTATCACCTCACATAGTTATC> 924
 HUMDAFC5 T.....A.....G.....CG.....A.A..C..C.....< 186
 HUMHBB T.....A.A.CCC.T.T.....< 65822
 HUMCDIR2 TA..CG.....G.....T..-..C.A..CC.....T.....CA> 1112
 HUMUKI3 T.....G.....AT..-..G..AC..CC.....TA..-..A..C..< 1943
 HUMAFP T.....T.....A.....G..< 19139
 GORAFPA T.....T.....A.....G..< 18491
 HUMTP004 T..C..T..CG...A.A..G..A.A..CCC.....> 986

MER33

HUMTFPB GGTAAATAAAATATATTATTAATAATTTACCTGTTCTTTTACTTTTCTAATGTGACTACTAGAAAACCTAAAA< 9070
 HUMFVIII12CG.....C.....-.....T.....CT.....C.....T.....< 477

Sequences used as probes are shaded.

MER25: HUMHBB (30772-31224) also not shown identical HUMHBEG (9524-9951); HUMBFXIII (1735-2173); HUMKM19 (446-152).
 MER26: HUMAMINONA (3-130); HUMAPOAI (157-41); HUMBFXIII (11660-11792); HUMIGLAMB (32880-33017); HUMXT00121 (155-286); HUMXT00273 (73-196).
 MER27: HUMAPB03 (2185-2131); HUMHBEG (1601-1484); HUMIL2RBA (79-236); HUMSIGMG3 (232-401).
 MER28: HUMHSSR (1635-1486); HUMMLC1G5 (645-243); HUMPHLAM (1486-1635); HUMTFPB;1 (8713-8992); HUMTFPB;2 (8216-8390).
 MER29: HUMCRYGBC (8014-8288); HUMGP91 (539-55); HUMPGEPI02 (3677-3936 and 4250-4485); HUMRSKP2 (269-573).
 'x' in HUMPGEPI02 indicates an Alu-Sx (10) insertion.
 MER30: HUMABLI9I(317-389); HUMABLIA2;1(87-217) and HUMABLIA2;2(470-522) the sequence fragments are separated by an Alu-Sb sequence indicated by 'x'; HUMCGPRA(2792-2889); HUMHPRTB(45702-45902); HUMINTB1A(733-524).
 MER31: HUMDMHI7(641-490); HUMSTXG2(4421-4580); HUMTATG5(220-382).
 MER32: GORAFPA (18363-18491); HUMAFP (19011-19139); HUMCDIR2 (1220-1112); HUMDAFC5 (47-186); HUMHBB (65961-65822); HUMPLG2BA (1077-924); HUMTP004 (898-986); HUMUKI3 (1795-1943).
 MER33: HUMTFPB (8990-9070); HUMFVIII12(337-477).

number based on 37 sequences may be realistic. Criticism can be made that GenBank is not a representative sample of the human genome. However, the number of Alu sequences estimated using the same assumptions is quite convincing. There are ~1800 Alu sequences in the GenBank release studied in this paper. This corresponds to ~600,000 Alu elements in the human genome. On the other hand, the experimentally determined numbers can be affected by two opposite factors: stringency and the quality of probes. The relatively low stringency ($2 \times \text{SSC}$ and 50°C) may lead to overestimates. However, the relatively high diversity of the MER repeats studied experimentally in this paper might have affected the results in the opposite direction.

DISCUSSION

Identification of new MER repeats is only a first step towards understanding their history, mutual relationships, and impact on the human (mammalian) genome. As computer-assisted research continues, we expect to detect less numerous and increasingly older families. Our paper indicates that we may be approaching limits of detectability of these repeats by DNA-DNA hybridization procedures.

Identification of novel repetitive families should give us more information about their rise and decline in the past. Some of the reported MER repeats may be distantly related to each other as they branched into subfamilies and diversified. One example may be homology between MER10 and THE-1 (39). These distant homologies can best be traced through multiple alignment of a large number of individual elements. Such studies will be fueled by the growing databases.

MER elements certainly affect genomic stability. In the previous paper one of the MER repeats (MER4) was shown to mediate multigene deletion on chromosome 14 (3). More examples of genetic rearrangements involving MER repeats have been shown in this paper. This illustrates the potential importance of MER repeats for the understanding of genomic variability associated with genetic diseases and evolutionary changes.

Finally, as indicated by recent reports (49,50), some of the most abundant repeats appear to be involved in gene regulation. It is likely that similar findings will be made for MER repeats.

ACKNOWLEDGEMENTS

We thank Drs Debbie Gemucio, Akitada Ichinose, Gerd Scherer, Anna Starzinski-Powitz, and Stuart Orkin for providing us with cloned DNA containing MER repeats, and Shyh-Wen Yang for his excellent computer assistance. This work was supported by grant DE-FG03-91ER61152 from the US Department of Energy, Human Genome Program (granted to J.J.) and by the State of Michigan Research Excellence Fund (J.S. and C.D.).

REFERENCES

- Jurka, J., Walichiewicz, J., and Milosavljevic, A. (1992) *J. Mol. Evol.* 35, 286–291.
- Zucchi, I., and Schlesinger, D. (1992) *Genomics* 12, 264–275.
- Jurka, J. (1990) *Nucleic Acids Res.* 18, 137–141.
- Kaplan D.J., Jurka, J., Solus, J.F., and Duncan, C.H. (1991) *Nucleic Acids Res.* 19, 4731–4738.
- Feinberg, A.P., and Vogelstein, B. (1983) *Anal. Biochem.* 132, 6–13.
- Southern, E.M. (1975) *J. Mol. Biol.* 98, 503–517.
- Benton, W.D., and Davis, R.W. (1977) *Science* 196, 180–182.
- Rogan, P.K., Pan, J., and Weissman, S.M. (1987) *Mol. Biol. Evol.* 4, 327–342.
- Gnarra, J.R., Otani, H., Wang, M.G., McBride, O.W., Sharon, M. and Leonard, W.J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* 87, 3440–3444.
- Jurka, J., and Milosavljevic, A. (1991) *J. Mol. Evol.* 32, 105–121.
- Labuda, D., and Striker (1989) *Nucleic Acids Res.* 17, 2477–2491.
- Skalnik, D.G., Strauss, E.C. and Orkin, S.H. (1991) *J. Biol. Chem.* 266, 16736–16744.
- Horn, G.T., Richards, B., Merrill, J.J., and Klinger, K.W. (1990) *Clin. Chem.* 36, 1614–1619.
- Ueda, S., Washio, K., and Kurosaki, K. (1990) *Genomics* 8, 7–12.
- Rotter, M., Zimmerman, K., Poustka, A. and Starzinski-Powitz, A. (1991) *Nucleic Acids Res.* 19, 1497–1504.
- Woods-Samuels, P., Kazazian, H.H., Jr. and Antonarakis, S.E. (1991) *Genomics* 10, 94–101.
- Chamberlain, J.S., Gibbs, R.A., Ranier, J.A., Nguyen, P.N., and Caskey, C.T. (1988) *Nucleic Acids Res.* 16, 11141–11156.
- Yen, P.H., Marsh, B., Allen, E., Tsai, S.P., Ellison, J., Connolly, L., Neiswanger, K. and Shapiro, L.J. (1988) *Cell* 55, 1123–1135.
- Collins, F.S., and Weissman, S.M. (1984) *Prog. Nucleic Acid Res. Mol. Biol.* 31, 315–462.
- Bottenu, R.E., Ichinose, A., and Davie, E.W. (1990) *Biochemistry* 29, 11195–11209.
- Shapiro, L.H., Ashmun, R.A., Roberts, W.M. and Look, A.T. (1991) *J. Biol. Chem.* 266, 11999–12007.
- Higuchi, K., Law, S.W., Hoeg, J.M., Schumacher, U.K., Meglin, N. and Brewer, H.B.Jr. (1988) *J. Biol. Chem.* 263, 18530–18536.
- Kimura, S., Hong, Y.-S., Kotani, T., Ohtaki, S. and Kikkawa, F. (1989) *Biochemistry* 28, 4481–4489.
- Vasicek, T.J. GenBank release 69.0.
- Adams, M.D., Kelley, J.M., Gocayne, J.D., Dubnick, M., Polymeropoulos, M.H., Xiao, H., Merrill, C.R., Wu, A., Olde, B., Moreno, R.F., Kerlavage, A.R., McCombie, W.R. and Venter, J.C. (1991) *Science* 252, 1651–1656.
- Ludwig, E.H., Blackhart, B.D., Pierotti, V.R., Caiati, L., Fortier, C., Knott, T., Scott, J., Mahley, R.W., Levy-Wilson, B. and McCarthy, B.J. (1987) *DNA* 6, 363–372.
- Samia, J.A., Alexander, S.J., Horton, K.W., Auron, P.E., Byers, M.G., Shows, T.B. and Webb, A.C. (1990) *Genomics* 6, 159–167.
- Akahori, Y., Handa, H., Imai, K., Abe, M., Kameyama, K., Hibiya, M., Yasui, H., Okamura, K., Naito, M., Matsuoka, H. and Kurosawa, Y. (1988) *Nucleic Acids Res.* 16, 9497–9511.
- Fujii, J., Zarain-Herzberg, A., Willard, H.F., Tada, M., and MacLennan, D.H. (1991) *J. Biol. Chem.* 266 (18), 11669–11675.
- Mackman, N., Morrissey, J.H., Fowler, B., and Edgington, T.S. (1989) *Biochemistry* 28, 1755–1762.
- Den Dunnen, J.T., Van Neck, J.W., Cremers, F.P., Lubsen, N.H. and Schoenmakers, J.G. (1990) *Gene* 87, 225–232.
- Stanford, D.R., Holicky, E.L., Perry, C.A., Rehder, K.Jr., Harvey, S.E. Rohleder, A.M. and Wieben, E.D. (1991) *DNA Seq.* 1, 357–363.
- Nomiyama, H., Tsuzuki, T., Wakasugi, S., Fukuda, M. and Shimada, K. (1984) *Nucleic Acids Res.* 12, 5225–5234.
- Chen, S.J., Chen, Z., d'Auriol, L., Le Coniat, M., Grausz, D., and Berger, D. (1989) *Oncogene* 4, 195–202.
- van der Feltz, M.J.M., Shivji, M.K.K., Allen, P.B., Heisterkamp, N., Groffen, J., and Wiedemann, L.M. (1989) *Nucleic Acid Res.* 17, 1–10.
- Pittler, S.J., Baehr, W., Wasmuth, J.J., McConnell, D.G., Champagne, M.S., van Tuinen, P., Ledbetter, D., and Davis, R.L. (1990) *Genomics* 6, 272–283.
- Edwards, A., Voss, H., Rice, P., Civitello, A., Stegemann, J., Schwager, C., Zimmermann, J., Erfle, H., Caskey, C.T., and Ansorge, W. (1990) *Genomics* 6, 593–608.
- Altruda, F., Cervella, P., Tarone, G., Botta, C., Balzac, F., Stefanuto, G., and Silengo, L. (1990) *Gene* 95, 261–266.
- Fields, C.A., Grady, D.L., and Moyzis, R.K. (1992) *Genomics* 12, 431–436.
- Rettenmeier, R., Natt, E., Zentgraf, H. and Scherer, G. (1990) *Nucleic Acids Res.* 18, 3853–3861.
- Ryan, S.C., Zielinski, R. and Dugaiczky, A. (1991) *Genomics* 9, 60–72.
- Gibbs, P.E., Zielinski, R., Boyd, C. and Dugaiczky, A. (1987) *Biochemistry* 26, 1332–1343.
- Calabi, F., Jarvis, J.M., Martin, L. and Milstein, C. (1989) *Eur. J. Immunol.* 19, 285–292.
- Ewulonu, K.U., Ravi, L. and Medof, M.E. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 4675–4679.
- Poncz, M., Schwartz, E., Ballantine, M. and Surrey, S. (1983) *J. Biol. Chem.* 258, 11599–11609.
- Loftus, J.C., Plow, E.F., Frelinger, A.L., D'Souza, S.E., Dixon, D., Lacy,

- J., Sorge, J. and Ginsberg, M.H. (1987) *Proc. Natl. Acad. Sci. U.S.A.* 84, 7114–7118.
47. Matera, G.A., Hellmann, U. and Schmid, C.W. (1990) *Mol. Cell. Biol.* 10, 5424–5432.
48. Batzer, M.A., Kilroy, G.E., Richard, P.E., Shaikh, T.H., Desselle, T.D., Hoppens, C.L. and Deiniger, P. (1990) *Nucleic Acids Res.* 18, 6793–6798.
49. Vidal, F., Mougneau, F., Gaichenhaus, N., Vaigot, P., Darmon, M. and Cuzin, F. (1993) *Proc. Natl. Acad. Sci. U.S.A.* 90, 208–212.
50. Brini, A.T., Lee, G.M. and Kinet, J.-P. (1993) *J. Biol. Chem.* 268, 1355–1361.